

## University of Groningen

### Cracking the code

Borleffs, Lotte Elisabeth

**IMPORTANT NOTE: You are advised to consult the publisher's version (publisher's PDF) if you wish to cite from it. Please check the document version below.**

#### *Document Version*

Publisher's PDF, also known as Version of record

#### *Publication date:*

2018

[Link to publication in University of Groningen/UMCG research database](#)

#### *Citation for published version (APA):*

Borleffs, L. E. (2018). *Cracking the code: Towards understanding, diagnosing and remediating dyslexia in Standard Indonesian*. [Thesis fully internal (DIV), University of Groningen]. Rijksuniversiteit Groningen.

#### **Copyright**

Other than for strictly personal use, it is not permitted to download or to forward/distribute the text or part of it without the consent of the author(s) and/or copyright holder(s), unless the work is under an open content license (like Creative Commons).

The publication may also be distributed here under the terms of Article 25fa of the Dutch Copyright Act, indicated by the "Taverne" license. More information can be found on the University of Groningen website: <https://www.rug.nl/library/open-access/self-archiving-pure/taverne-amendment>.

#### **Take-down policy**

If you believe that this document breaches copyright please contact us providing details, and we will remove access to the work immediately and investigate your claim.

*Downloaded from the University of Groningen/UMCG research database (Pure): <http://www.rug.nl/research/portal>. For technical reasons the number of authors shown on this cover page is limited to 10 maximum.*

# CHAPTER 8

---

General discussion

---

## 8.1 INTRODUCTION

The ability to read is an essential skill in modern societies. Most children become proficient readers if they are provided with adequate learning opportunities, have no significant sensory deficit, and if their cognitive capacity is not severely compromised (Richardson & Lyytinen, 2014). Unfortunately, however, this does not apply to all. Dyslexia is the most common learning disability (Cortiella & Horowitz, 2014) and occurs in all languages (Shaywitz et al., 2008), even though the specific orthography to be acquired has been identified as having a great impact on reading acquisition and dyslexia (Ziegler & Goswami, 2005; Seymour et al., 2003). Dyslexic students face persisting problems with reading and spelling, with possible negative effects on their cognitive development, school motivation, well-being, and self-esteem (Lovio et al., 2012). Early identification of those students who are at risk of developing serious reading deficits and the provision of effective support are therefore of great importance to break the vicious cycle of negative learning experiences and to minimize the sequelae of developing or ongoing reading disabilities.

The past decades have seen progress being made in our understanding of typical reading development and the causes of deficits in the acquisition process. Research on reading and writing, however, has traditionally focused on a limited number of European languages, in particular English, a language with an exceptionally inconsistent and irregular orthography (Share, 2008). More recently, research on reading and spelling in other languages and scripts has been receiving increased attention (Landerl et al., 2013; Moll et al., 2014; Peterson & Pennington, 2015; Winskel, 2013). Yet, whereas the body of research focusing on East Asian languages is growing (i.e. Chinese, Japanese and Korean), still very little research has been conducted on reading and spelling development in languages of Southeast Asia, among which is the highly transparent Standard Indonesian (SI) language.

To the best of our knowledge, prior to this PhD research project, neither had there been developed a standardized assessment battery to identify students with or at risk of developing dyslexia, nor a standardized intervention scheme to support struggling readers in SI. The general aims of this thesis were therefore 1) to gain more insight into orthographic differences and their impact on reading and dyslexia, 2) to compose a test battery to facilitate the assessment of reading difficulties in young readers of SI, and 3) to develop an SI version of the computerized-reading intervention tool GraphoGame, and test its effectiveness in first grade readers of SI.

In what follows, the results of the previous chapters are summarized. Main findings are discussed in relation to existing literature, and future perspectives are addressed.

## 8.2 ORTHOGRAPHIC DIFFERENCES AND THEIR IMPACT ON READING DEVELOPMENT

In Chapter 2, we discussed the orthographic transparency, syllabic complexity, and morphological complexity of alphabetic languages in relation to reading acquisition and dyslexia. Following that, in Chapter 3, we reviewed various metrics proposed in the literature for quantifying these three concepts in alphabetic languages.

At a general level of description, all researchers agree that the basic processes of reading are the same for all languages, for instance in terms of matching inputs to memory, association, retrieval, decomposition, decoding, and assembly. Above and beyond these universals, however, important differences do emerge at more detailed levels. The way in which phonological, orthographic, and morphological processes function, is shaped by the specific orthography being used, necessitating orthographic-specific strategies when learning to read. With improving skills, the reader increasingly manages to adapt reading procedures to the demands of the writing system to thus improve reading efficiency. These specific and vital adjustments that depend on the orthography used, have implications for new readers and the development of reading difficulties like dyslexia. Despite our growing insight in these processes we still need to learn more about orthographic differences and about how to take advantage of language-specific orthographic, syllabic, and morphological sources of information.

Considering the above, detailed knowledge of differences between orthographies and metrics to measure these differences can provide a stepping stone in the development of language-specific reading instructions and interventions. The development of reliable metrics to measure differences between language scripts, however, hasn't received much attention so far. Based on the review presented in Chapter 3, we concluded that although the predictive value of these metrics is promising, more research is needed to validate the value of the metrics discussed and to understand the 'developmental footprint' of orthographic transparency, morphological complexity and syllabic complexity in the lexical organization and processing strategies being developed. One general difficulty was the variety in definitions of orthographic transparency, morphological complexity and syllabic complexity used. To make predictions about whether, and if so how, any of these orthographic aspects might affect reading acquisition and skilled reading, one cannot go without a clear and widely accepted definition of the specific aspect studied.

Regardless of such limitations, the studies discussed in both Chapter 2 and 3 do trigger our thoughts about the complexity or simplicity of languages. There is general consensus about the approximate classification of several languages in terms of their orthographic transparency (e.g. Seymour et al., 2003). Considering orthographic transparency as a continuum, one can be certain about its extreme

positions (e.g. the regular Finnish orthography at one extreme and the irregular English orthography at the other), even though the objective location of each orthography on this transparency continuum may remain uncertain (Aro, 2004). Studies have suggested that with regard to reading acquisition, transparent orthographies with high grapheme-phoneme consistency are more easily acquired than opaque and complex writing systems featuring a large number of inconsistent and irregular spellings (Aro & Wimmer, 2003; Seymour et al., 2003). From a transparency point of view, reading fluency in Finnish would hence be easier to acquire than in English. When it comes to word-recognition, however, characteristics of the Finnish morphology reduce the effectiveness of these 'beneficial' factors as the majority of Finnish words are polysyllabic and tend to be long due to the highly productive compounding, a rich derivational system, and agglutinative morphology (Aro, 2004; Lyytinen et al., 2006). English scores the lowest on the morphological complexity measures type-token ratio (TTR), moving-average type-token ratio (MATTR), and Juola method (Kettunen, 2014) among all languages included, whereas Finnish was found to be the most (or second most) morphologically complex language. Nonetheless, behavioural research has shown that more than 95% of Finnish students acquires accurate reading skills during the first year of reading instruction (Holopainen et al., 2001), while the rate of early reading acquisition was suggested to be slower by a ratio of about 2.5:1 in English than in most European orthographies (Seymour et al., 2003).

In opaque orthographies such as English or Danish, the mastery of the alphabetic principle provides only part of the key for decoding and many words cannot be sounded out accurately without having access to the stored phonological representation of the whole word. Some researchers have hypothesized that there is a threshold which, once exceeded, results in an abrupt change in the way foundation literacy is acquired; according to Seymour et al. (2003) reading acquisition in the deeper orthographies (e.g. Portuguese, French, Danish, English) may be based on the formation of a dual (a logographic and an alphabetic) foundation, which takes more than twice as long to attain as the single (alphabetic) process when learning a shallow orthography (e.g. Dutch, Finnish, German, Greek, Icelandic, Italian, Norwegian, Spanish, and Swedish). Rather than focusing on two separate processing routes, others have focused on the different sizes of the orthographic units the reader uses during reading and decoding (Ziegler & Goswami, 2005). Ziegler and Goswami hypothesize that in orthographically consistent alphabetic languages children rely heavily on grapheme-phoneme recoding strategies as these correspondences are relatively consistent, whereas in less consistent orthographies, children cannot use smaller grapheme units as easily because smaller grain sizes tend to be less consistent than larger grain sizes (Treiman et al., 1995). This may well lead to the development of multiple decoding

strategies that enable the learner to decode at several different grain sizes, supplementing grapheme-phoneme correspondences with the recognition of letter patterns for rimes and attempts at whole-word recognition (Ziegler & Goswami, 2005), which demands the engagement of a wider range of cognitive skills. Interestingly, although variation in orthographic depth may indeed result in differences in the rate of reading development, several researchers have argued that still the same cognitive mechanisms are underlying reading acquisition across different languages (Caravolas et al., 2013; Perfetti & Harris, 2013; Vaessen et al., 2010).

In languages in which the morphological structure of a given word hardly ever changes depending on its function in the sentence or the phrase it belongs to, a word that has been stored in the lexicon will be retrieved with little effort (Acha et al., 2010). In contrast, the agglutinative nature of the Finnish (or Turkish, Basque) morphological system results in words of considerable length that contain multiple parts of the semantic information. Yet, more than one third of Finnish children is able to read before the start of formal reading instruction (Holopainen et al., 2001). This suggests that Finnish children acquire efficient strategies to overcome the potential difficulty resulting from the morphological complexity of the Finnish language. Finnish children have been argued to be highly oriented toward the details of spoken language in order for them to differentiate words with small (single phonemic) variations. This would account for the large number of exceptional inflections that are already understood by Finnish children at school-entry age (Torppa et al., 2010). Whereas Finnish children use the grapheme–phoneme correspondences to read Finnish words during their first school year (Holopainen, Ahonen, & Lyytinen, 2002), in the third year they have been shown to read morphologically complex words better and faster than mono-morphemic words, especially if the words are low-frequency words (Bertram, Laine, & Virkkala, 2000). These findings suggest that children learn to recognize Finnish words through the recognition of their constituent stems and morphemes (Acha et al., 2010). It has even been suggested that only very high-frequency inflected words may be stored as whole word representations while the majority of inflected Finnish words seem to be represented in decomposed form (Vartiainen et al., 2009). Adult readers of Finnish have been argued to switch between the use of lexical and phonological processing strategies, both within and between words depending on the familiarity, structure and length of the word. In dyslexic readers, however, these phonological and orthographic processes seem to be less tightly integrated, resulting in a less effective predominant reliance on one or the other processing mode (Leinonen et al., 2001). Still, reading fluency rather than accuracy is seen as the most central factor being compromised among dyslexic readers in transparent orthographies such as Finnish (Lyytinen et al., 2015).

With regard to the syllabic complexity measures, the results of the data-driven automatic syllabification algorithm (SbA) by Adsett and Marchand (2010) were in line with previous work applying a structural (Ramus et al., 1999) and behavioural approach (Seymour et al., 2003), and resulted in similar distinctions between orthographies. Using their SbA-algorithm, Adsett and Marchand ranked languages based on the difficulty to automatically syllabify the language, whereby an increased difficulty to syllabify signified increased syllabic complexity. Ramus et al. (1999) used the duration of vocalic and consonantal intervals as indicators of syllabic structure and complexity (also referred to as *timing*; Aravaniti & Rodriguez, 2013), whereas Seymour et al. (2003) investigated the number of open versus closed syllables and consonant clusters. When comparing Adsett and Marchand's SbA-pronunciation results for Dutch, English, French, Italian, and Spanish with the speech results Ramus et al. (1999) had obtained for these languages, lower word accuracies were obtained in the SbA-pronunciation domain (feedforward direction) for the languages judged by Ramus et al. to have a more complex syllabic structure (Dutch and English) than those believed to be syllabically less complex (French, Italian, and Spanish). Word accuracy was quantified as the number of words syllabified by the SbA-method in exactly the same way as was given in the lexicon for that language. Seymour et al. recorded significantly lower error rates during pseudoword reading in first- and second-grade children whose native language was Finnish, French, Greek, Italian, Spanish, or Portuguese (all languages that Seymour et al. perceived as having simple syllable structures), compared to error rates for the children learning to read languages they considered to be syllabically more complex (Austrian, Danish, Dutch, English, German, Icelandic, Norwegian, and Swedish).

Syllabic complexity is thought to affect how readily children become sensitive to the phonological structure of language (Duncan et al., 2006). Children who speak French, a language regarded as having a relatively simple syllabic structure characterized by a predominance of open syllables, were found to demonstrate more phonological awareness prior to any formal instruction than their counterparts speaking the syllabically more complex English language (Duncan et al., 2006). Several studies support the hypothesis that specifically consonant clusters may pose an additional problem to the young learner (e.g. Lee & Wheldall, 2011; Seymour et al., 2003; Treiman & Weatherston, 1992). The embedding of grapheme-phoneme correspondences in consonant clusters has even been suggested to impede the reading acquisition process (Seymour et al., 2003). Correspondences between the grapheme 'p' and phoneme /p/ in English, for example, occur in isolation but also in various consonant clusters such as 'sp', 'spr', and 'mple'. Arguably, in languages with a greater syllabic complexity, material for new readers will necessarily require more skill in recognizing such correspondences in clusters, slowing down the learning process. Clusters are possibly treated as phonological units and are difficult

to split into phonemes (Treiman, 1991). The high level of co-articulation in the consonant phonemes in the cluster might exacerbate the problem (Serrano & Defior, 2012).

For future research, we would suggest that more cross-linguistic studies be conducted comparing two orthographies which are similar on as many aspects as possible, but different on the particular component of interest. Furthermore, within-language studies (e.g. using vowelized and unvowelized scripts in Hebrew) that are capable of isolating a particular aspect that has been argued to drive cross-linguistic differences may also provide valuable information. The measures proposed in Chapter 3 may be used to compare languages on the specific aspect investigated or may provide a starting point for other research focusing on the development of quantitative measures. Knowing that this will be a difficult task, various studies will need to be conducted on the same set of orthographies, and these studies will need to be replicated in other languages. Moreover, the proposed complexity rankings will need to be supported by behavioural data of reading acquisition and skilled reading to validate their value in the study of reading development and to teach us more about techniques to help struggling readers to take advantage of language specific orthographic, syllabic, or morphological resources.

## 8.3 SCREENING AND ASSESSMENT OF CHILDREN LEARNING TO READ IN SI

### **8.3.1 Test battery and diagnostic criteria**

In Chapter 4, we described the development of a test battery to facilitate the assessment of reading acquisition in SI and the early detection of reading difficulties. Moreover, we presented the first data obtained with this battery among 139 first- and second-graders recruited from elementary schools in Jakarta (Java), and proposed preliminary criteria for the categorization of beginner-readers of SI as ‘typical readers’ or readers ‘at risk of dyslexia’ based on the outcomes for reading and decoding fluency and spelling. In Chapter 5, we further analysed the Jakarta data, and combined those data with the test results of an additional sample of 146 second- and third-graders recruited in Medan (Sumatra), with the aim to investigate profiles of cognitive predictors of reading found among young readers of SI, and to test the fit of single versus multiple deficit models of dyslexia to individuals categorized as ‘typical readers’ and ‘at risk of dyslexia’ by using part of the in Chapter 4 described test battery.

The initial test battery consisted of nine reading and reading-related tasks assessing word reading, pseudoword reading, arithmetic, rapid automatized naming, phoneme deletion, forward and backward digit span, verbal fluency, and passive (orthographic choice test) and active spelling (writing to dictation). A selection of these tests was also used in Chapters 5 to 7, with the inclusion of an additional



auditory synthesis task and two subtasks of the SON-R in Chapters 6 and 7. The need for future research to further investigate the reliability of our measures, to standardize the preliminary norm scores, and to fine-tune our at-risk criteria has been stipulated more than once in the past chapters. Strong correlations between reading fluency, decoding fluency, and the reading related skills, in addition to a reoccurring factor structure in the four subsamples described in Chapter 5 (Table 5.3; see Chapter 4, Table 4.4, for additional factor analyses conducted on the Jakarta sample), however, have so far provided support for the reliability of our test battery. To collect more generalizable normative data, large scale assessments including different age groups need to be conducted in other parts of the country including a larger variety of ethnic groups and different socio-economic backgrounds than only those included in the Jakarta and Medan samples.

Based on our categorization criteria selecting the lowest 10% of readers, decoders and/or spellers, 17% of the first graders, 14-18% of the second graders and 15% of the third graders were categorized as being at risk of dyslexia (see Table 5.4). These numbers are relatively high compared to the general prevalence rates found in other languages. Prevalence rates vary according to definitions and diagnostic criteria, but are generally assumed to lie between 5-10% for alphabetic orthographies, with relatively low rates occurring in the more transparent orthographies, and with numbers exceeding 10% for the exceptionally opaque English orthography (Habib & Giraud, 2013; Shaywitz, 1998). It seems therefore unlikely that all students that were labelled as 'at risk of dyslexia' in our SI sample are indeed dyslexic or will develop dyslexia in the future.

A possible explanation for these high rates could be that the reading tests developed were slightly too easy considering the relatively high reading levels in our sample (see Table 5.5), which may have reduced the discriminatory power of our tasks. High accuracy levels already at the end of first grade are not uncommon in transparent orthographies such as SI. Reading accuracy for pseudowords was already around 85% at the end of first grade for the German, Dutch, Spanish, and Finnish children and over 90% for the Swedish children included in Aro and Wimmer's study (2003). By contrast, the English students had achieved a 50% accuracy level only. Moreover, a relatively small number of students that were categorized as 'at risk' across grades based on their reading and/or decoding scores (i.e. 11 out of 43 in Table 5.4), fell within the lowest 10% on both reading and decoding tasks, whereas the remaining 32 students were classified as 'at risk' based on one of the tasks only, with the other score being above 10% (and below percentile 40). This may have resulted in the overall 'at risk' rate of 16%, well above 10%. In the combined sample, reading and decoding fluency correlated at .751. When looking at individuals grades, we found correlations of .842 (Jakarta grade 1), .566 (Jakarta grade 2), .697 (Medan grade 2), and .740 (Medan grade 3). Higher

correlations could have resulted in more overlap between categories and hence in lower at risk rates.

One possible solution for the relatively high at-risk prevalence rates in our sample would be to adjust our criteria. Reading scores of 1 to 1.5 $SD$  below grade level are widely accepted in research and clinical practice as standard criteria for dyslexia (Landerl et al., 2013), and an increase of the current 1.28 $SD$  to 1.5 $SD$  below the subsample's mean would select the lowest 6-7% of the scores instead of the lowest 10%. This, however, does not have our preference. As our assessment battery has only recently been developed, a first step would be to optimize our reading and spelling tests to further augment their reliability and discriminatory potential. Our word reading task was modelled on an existing one minute reading task (i.e. EMT, Brus & Voeten, 1979) with its content being drawn from commonly used first grade Indonesian textbooks. The words included ranged from one to four syllables in length (e.g. *es* 'ice'; *mata* 'eye'; *menyenangkan* 'pleasant'). As a first adjustment we would suggest to increase the length and complexity of the reading fluency task by turning our one minute reading task into a three minute task, in which the students would have to read three times for a duration of one minute (similar to the Dutch DMT [Three Minute Test], CITO). Sections one to three of this new test would need to increase in complexity, by adding more complex, inflected, infrequent words taken from higher grades' textbooks and those words containing less commonly used syllable structures and consonant clusters. With more items and a larger variety in complexity of the items, the reliability of the test and the normality of the score-distributions increase. As noted in the method section of Chapter 5, decoding fluency was defined as the number of pseudowords read in one (Jakarta) and two (Medan) minutes. As can be seen in Table 5.4, the number of children with a score below 10% in the two-minute administration approximates the theoretical distribution (observed 8 and 8, expected 7.4 and 7.2 in grade 2 and 3 respectively) slightly better than the one-minute administration (observed 5 and 6, expected 7.5 and 6.4 in grade 1 and 2 respectively). This shows that extending a test could potentially improve the distributional characteristics of the test scores. Moreover, it would also be interesting to see whether the addition of a reading text (to be read silently or aloud) would generate further useful information.

With regard to spelling, the mean task scores across grades on the orthographic choice test (OCT) and writing to dictation discussed in Chapter 5 came close to these tasks' absolute maximum score (see Table 5.5), especially among typical readers, even though task complexity of the spelling tasks had already been increased for the Medan sample by adding longer words taken from second and third grade textbooks. Only three out of 46 children in our combined Jakarta and Medan sample described in Chapter 5 met the at-risk criteria solely based on spelling and orthographic knowledge (see Table 5.4). Nonetheless, we believe that there are enough grounds

to keep spelling as part of our current criteria. Several studies, including some conducted in transparent orthographies, have demonstrated that children with reading difficulties are often poor in both reading and spelling (De Jong & Van der Leij, 2003; Eklund et al., 2015; Pennington & Lefly, 2001; Puolakanaho et al., 2008; Van Bergen et al., 2012). In line with that we found the three students referred to above to be poor spellers, as well as relatively poor readers (scores below the 18th percentile on reading fluency). Moreover, about a third to a quarter of the children with severe reading problems in our sample also had severe difficulties in spelling and/or orthographic knowledge (see Table 5.4). Spelling, in addition to reading, therefore plays an important role in determining a targeted treatment plan after the dyslexia diagnosis has been confirmed.

Another linguistic aspect that might have influenced reading performance, is that monosyllabic words are uncommon, and that Indonesian children are therefore trained to read long words from an early age as instructions in primary-school books already contain multisyllabic words. Two-, three-, and even four-syllabic words may therefore have posed relatively little problem to the readers in our sample. As alluded to above, children learning to read in the agglutinative but also highly transparent Finnish orthography, have been suggested to be highly oriented toward the details of spoken language to be able to differentiate words with small variations (Torppa et al., 2010). SI also possesses a rich system of morphemes and affixations which differ depending on the word class of the stem (Winksel & Widjaja, 2007). Interestingly, however, colloquial spoken SI often uses non-affixed forms. Hence different from Finnish, one would expect Indonesian children to have less exposure to these longer multisyllabic words in spoken form to support their reading. Unfortunately, no numbers are available for the percentage of Indonesian children acquiring reading skills within the first school year, which would be interesting to compare to other highly transparent languages.

### **8.3.2 Predictors of reading skills**

As previously mentioned, in Chapter 5 we assessed reading, spelling, phonological skills, and non-verbal IQ assessed in a total of 285 first-, second-, and third-graders, to investigate which profiles of cognitive predictors of reading were found among young readers of SI classified as being at risk of dyslexia based on the criteria and tests proposed in Chapter 4. For each individual case, first the presence of deficits in phonological awareness (PA), verbal working memory (VWM) and naming speed (NS) was determined using a 10<sup>th</sup> percentile cut-off, after which regression analyses were conducted to test the fit of single- and multiple-deficit models of dyslexia to individual cases.

Accounting for 33% of the cases that satisfied both methods of individual prediction, Pennington et al.'s (2012) Hybrid Model proved the best fit, proposing

that there are multiple pathways to being at risk of dyslexia in SI, some involving single deficits and some multiple deficits. None of the deficits in PA, NS, or VWM alone was sufficient to predict a risk of dyslexia in our SI sample, nor was a deficit in PA necessary. Whereas PA was the strongest single predictor in Pennington et al.'s samples, NS was the main predictor of reading and decoding fluency in our sample, followed by PA and VWM. PA was the most common deficit in our sample.

Fifty-four percent of the children (25/46) we classified as at risk of dyslexia had one or more deficits, compared to 68-78% in Pennington et al.'s sample. A possible explanation for the lower numbers obtained could be that our test results might have had a higher "noise" level when compared to the testing conducted by Pennington et al., who used highly optimized tasks which have already been used in previous studies. Interestingly, however, still 22-32% of dyslexic cases in Pennington et al.'s study did not have any deficit based on a 10<sup>th</sup> percentile cut-off in PA, language skill or NS/processing speed.

In line with the conclusion drawn in Pennington et al.'s paper for English and Norwegian, our results indicate that the relation between predictors, deficit profiles and reading problems in SI based on our assessment battery are probabilistic and not deterministic. These findings raise questions about the predictive value and clinical utility of these deficit profiles in the diagnostic process involving dyslexia, as they imply that students with unexpected (despite adequate learning opportunities, without significant sensory deficits or severely compromised cognitive abilities) and persistent reading and spelling difficulties in SI, do not necessarily need to fit a particular deficit profile or even need to have any deficit in PA, NS, or VWM at all. Still, we think that assessing these cognitive skills can provide valuable and converging information to support the dyslexia diagnosis in SI.

Although there does not seem to exist one typical at risk of dyslexia profile in our SI sample, it is worth noting that out of 21 at-risk cases without deficits, 15 had borderline scores (percentile 11 to 25) on one or more of the predictors included. Forty out of 46 readers being categorized as at risk of dyslexia in our sample (87%), had scores on PA, NA and/or VWM within the lowest 25% of their grade. At-risk students without at least borderline deficits on these reading-related cognitive skills were therefore more uncommon. Moreover, we recommend to always include both individual teacher ratings and school grades in the final diagnosis, as our results showed significant differences between the typical and the at-risk-of-dyslexia readers on both variables that supported the test results of the reading assessment battery (see Table 4.6).

Pennington et al. found no significant differences in relative importance of predictors when comparing English with the more transparent Norwegian script among kindergarten and first grade students. By contrast, Furnes and Samuelsson (2010), who drew a sample from the same international twin study used by

Pennington et al., reached different conclusions for their kindergarten through second grade sample; whereas PA as a predictor of reading skill in the Scandinavian (Norway, Sweden) sample was limited to the end of first grade, it remained a significant predictor in the two English-speaking samples (Australia, United States). NS was similarly predictive of reading at first and second grade across orthographies. When analysing our data further per subsample with stepwise multiple linear regression (see Discussion Chapter 5), we found a similar pattern for grades 1 and 2 in Jakarta. While PA and NS were both significant predictors of reading and decoding fluency in first grade Jakarta, only NS remained a significant predictor in second grade. These findings agree with results from earlier research in languages with transparent orthographies and indicate a decreasing effect of phonological awareness on reading after starting school (formal instruction) when the basic decoding rules have been learned (e.g. De Jong & Van der Leij, 2002; Georgiou, et al., 2008; Holopainen et al., 2001), whereas the importance of NS over time appears to increase (De Jong & Van der Leij, 1999; Vaessen & Blomert, 2010). It should be noted, however, that both PA and NS remained significant predictors of reading and decoding fluency in grades 2 and 3 in Medan.

Although there is a general consensus about NS, PA and VWM being predictors of reading skills, the differences between studies in general on the relative weight of these predictors of reading highlights the complex and multifactorial character of this developmental problem; different results may relate to the transparency of the orthography, but also to the developmental phase of reading of the participants included, the type of measures used, and even to the definitions and inclusion criteria employed. Another interesting possibility, however, could be that we just haven't found *the* predictor of reading skills yet, or that the specific skill needed for reading and decoding fluency can only be tested by assessing reading and decoding themselves.

In the case of dyslexia, Aravena, Tijms, Snellings and Van der Molen (2017) argue for a combination of conventional testing and the use of new measures. The authors argue that a key factor in dyslexia remains a fundamental disability to learn letter-speech correspondences, and that the manifestation of this deficit can be evoked by presenting them with a novel script (also see Aravena, Snellings, Tijms, & Van der Molen, 2013). In order to do so, the researchers assessed letter-speech sound learning using an artificial script to differentiate between dyslexic and normal readers of Dutch. A benefit of using an artificial script was that it allowed for controlling for differences in prior exposure. Moreover, the authors hoped that with the focus on learning rather than on the level of skill already obtained, they would be able to capture learning in action and identify factors that interfered with the learning of grapheme-phoneme correspondences. Their dynamic assessment aimed at training eight basic letter-speech sound correspondences and was followed by a

short test assessing these correspondences in addition to the student's word reading ability of this unfamiliar script. Their findings indicated that dyslexic readers were outperformed by typical readers both on the artificial script's letter-speech sound identification task (speed and accuracy), and on the word reading task. Aravena and colleagues (2017) analysed furthermore the relative contribution of their new measures in relation to the more conventional predictors PA and RAN. Their results showed that the artificial orthography related predictors made meaningful and partly independent contributions to the explanation of individual differences in reading and spelling abilities. The researchers therefore argue that a combination of conventional testing with the new measures proposed would result in a stronger predictor of individual differences in reading and spelling abilities compared to conventional testing alone, showing the potential benefit of this procedure for the assessment of dyslexia in clinical practice.

#### 8.4 GRAPHOGAME SI

In Chapter 6, we described the theoretical background, the development, and design of our SI edition of GraphoGame. Our aim was to discuss general GraphoGame principles, and to elucidate some of the specific choices we made for our SI version, hoping that our study will be an inspiration for the future development of additional language versions of this or similar digital-based learning environments. Two versions of GraphoGame for SI were developed: the main design that included 21 'streams' subdivided into 333 'levels,' and a compressed version offering the same 21 streams in 177 levels. In line with other GraphoGame studies (e.g. Kyle et al., 2013; Saine et al., 2010), the main game design was aimed at five 10-15 minute playing sessions per week (also see Richardson & Lyytinen, 2014). Moreover, the shorter version was created to anticipate school settings in which this preferred playing frequency would not be possible due to practical restrictions and where the complexity of the game content still needed to coincide with the level of the students' regular classroom reading instruction.

To effectively support children during the early stages of reading acquisition, the GraphoGame method makes use of the most functional sublexical units of the particular orthography: in order to help the young reader match the mappings of the orthography to the spoken language form, the training focuses on the most frequently and (in less transparent orthographies) most consistently used correspondences between the smallest distinguishable parts specific to the language being learned (Richardson & Lyytinen, 2014). Research has shown that when teaching reading in transparent orthographies with consistent grapheme-phoneme correspondences, it makes sense to focus on exactly these connections (Holopainen et al., 2002; Landerl, 2000). To typically developing readers, this approach is the quickest and simplest way to learn to decode. Moreover, the game

stimulates the child to use sublexical units, both graphemes and syllables, as building blocks to learn to read and spell whole words.

Besides a description of how and why we developed our version of GraphoGame the way we did, in Chapter 6 we additionally discussed the results of the pilot study among 69 typical and struggling beginning readers recruited from first-grade classrooms of an elementary school in the city of Medan. Between October and March, the students attended on average nine (range 5-12) GraphoGame sessions of 15-20 minutes. Although large-scale randomized controlled studies are needed to confirm the effectiveness of our SI reading game, the GraphoGame player data collected during this first pilot study showed that most of its variables were significantly related to mid- and post-test reading and decoding skills. Moreover, the results indicated that the more the students with low pre-test phonological skills were exposed to the game, the better their post-test performance on reading and decoding fluency became.

Considering the fact that the main game design would merely offer players more opportunities to practice the same game content, we questioned the added benefit of extended training using the same grapheme-phoneme correspondences, syllables and words for those students who were able to attain decoding fluency with regular classroom instruction and the short version played 1-2 hours spread out over several weeks and sessions. Once the grapheme-phoneme correspondences have been learned, word decoding in transparent orthographies such as SI is expected to be attained by basically putting together the different sounds of written words (Richardson & Lyytinen, 2014). Decoding skills are then automatized after extensive reading exposure and repetition, and for those good players, the compressed game design may already provide sufficient practice. For the more struggling players, we argued that the extended main game design could be beneficial as hopefully even more practice and game exposure would result in further improved reading and decoding skills.

In Chapter 7, we tested the main game design during a more intensive training by having 33 first-graders from the outskirts of Medan, play the full game more frequently during a shorter period of time (aimed at five 10-15 minute playing sessions a week for 13 weeks). The aims of this second pilot study were to further evaluate GraphoGame SI's usability, test the effectiveness of our main game design in promoting reading (-related) skills in first-grade learners of SI, and use the data gained during this study to further improve the game design. Our correlation results show that progress in the game (as measured as highest game level reached, the total number of levels played, and the average number of levels played per minute) and several of the reading-related skills we assessed were significantly related to reading and decoding fluency skills. Moreover, the highest level reached was found to be a significant predictor of reading and decoding abilities both at the post-test

and the 4-month follow-up assessment. Our results additionally indicated a significant interaction effect: only for students with average to above-average pre-test letter knowledge, progress in the game was strongly related to reading and decoding fluency at post-test. No such significant difference was found between high and low game progress in students with below average letter-sound knowledge at pre-test. If either one of the two stayed behind (i.e. low letter-sound knowledge or low game progress; see Figures 7.1 and 7.2), reading and decoding fluency at post-test were below average. This suggests that there is an optimal stage at which GraphoGame is effective. As a first step, in order to hopefully enable students with low letter knowledge at the outset to also fully benefit from the game, we proposed an extension of the playing period to approximately six months. For those students, thirteen weeks of gameplay may have been too short to firmly establish their letter-sound knowledge and to build up their phonological awareness skills (see Glatz et al., submitted, for a similar view).

This second pilot study evaluating the main GraphoGame SI design further validates our digital reading intervention by showing once more that there is a relationship between game progress and reading and decoding proficiency as tested shortly after the last playing session. Based on our preliminary but promising results, the next step would be to implement GraphoGame SI in a larger, more diverse sample of beginning readers and include active and passive control groups to further investigate and improve its effectiveness in promoting basic reading and reading-related skills. Moreover, it may be worthwhile to also assess motivation, familiarity with the use of a computer (even though a short training was provided to all students), familiarity with the type of assessment used, or other (reading-related) classroom activities to investigate their potential impact on the results. Another aspect worth investigating would be whether the tool might have potential as a diagnostic instrument in the early identification of readers at risk of developing serious reading deficits or dyslexia while simultaneously offering additional learning opportunities. Regular paper-and-pencil tests may not be sensitive enough to detect subtle differences within brief periods of time. Therefore, changes in reading related skills may be better captured by data provided by the game itself. By incorporating separate "proficiency" games in-between the regular training levels, one could measure improvement in accuracy and response times for individual learners. Paper-and-pencil tests assessing letter knowledge, for instance, quickly reach ceiling levels and thus lose their predictive power, whereas the speed of grapheme-phoneme associations may still provide valuable information (also see Glatz et al., submitted). As discussed in Chapter 7, the in-game assessments letter-sound knowledge I and II and the highest GraphoGame level reached, correlated strongly with post-test reading and decoding fluency (see Table 7.4 and 7.5a). Moreover, letter-sound knowledge I and GG highest level were shown to be significant



predictors of the latter two skills (see Table 7.6). Speed of letter-speech sound associations was not included yet in the current GraphoGame SI in-game assessments, but could easily be incorporated in future versions. These results underline the potential of GraphoGame SI as a dynamic assessment tool.

Future research will need to uncover why some students did and others did not seem to benefit sufficiently from the game and which adjustments are necessary to increase GraphoGame SI's effectiveness. Further improvement of the stimulus selection algorithm will hopefully at some point make the compressed version of this game redundant and will enable the game to truly adapt to the student's abilities. Another aspect worth looking into would be the fact that our GraphoGame version does not (yet) train the ability to sound out graphemes, syllables, and words. The way the game content is currently presented, the students are not instructed to read any of the targets out loud. Without complicating the game too much with the incorporation of sound recognition software, future research could start by simply asking the students to repeat the auditory stimulus (i.e. phoneme, spoken syllable or word) before choosing the corresponding written item on the screen. Besides keeping the children focused on the task, it could potentially also further strengthen the learning of the correct correspondences of spoken and written forms. Moreover, a flashed presentation (e.g. at 2000, 1000, 500, or 200 ms) of the words trained in previous levels could be added to the game design to further promote reading fluency, similar to one of the subtasks of the Dutch reading intervention program Build! (Regtvoort, Zijlstra, & Van der Leij, 2013).

In addition to the relationship between phonology and orthography, current models of children's reading development emphasize the importance of semantics (e.g. Ehri, 2005). Specific knowledge of the target words has been shown to predict reading accuracy in beginner readers (e.g. Duff & Dulme, 2012). In their 2015 paper, Van Gorp, Segers, and Verhoeven present a word identification game in which poor-reading second-grade learners of Dutch were asked to make several types of semantic categorizations while simultaneously practicing word identification, with the aim to enhance their decoding efficiency (i.e. combined reading speed and accuracy). During the game, the players were asked to categorize a target word presented on the screen (e.g. *hamster*, 'hamster') as quickly and accurately as possible by dragging and dropping the word into the correct semantic category (e.g. 'animals' versus 'garbage/ pseudowords'). Van Gorp et al.'s study showed significant increases in decoding efficiency of the players when compared to the waiting list control group. The authors concluded that their relatively short intervention game (five hours of playing across a period of five weeks) elicited transfer and retention effects when assessed using a standardized read-aloud test containing lists of untrained words and pseudowords. Unlike word decoding where the reader only processes orthographic and phonological information, word identification requires

the reader to also actively address semantics during the reading process (Van Gorp et al., 2015). The addition of a word identification task using semantic categorization would be an interesting aspect to further investigate for SI and to be added to our next version of GraphoGame SI.

To conclude: The mechanisms of typical reading acquisition and the causes of deficits in this development remain complex, which makes it all the more fascinating that this process comes so naturally to many of us despite differences in socio-economic backgrounds, intellectual capacities, and the characteristics of the language being learned. In 2016, according to UNESCO, less than 1% of Indonesians between 15 and 24 years old could not read compared to 30% of adults of 65 years and above. Hence tremendous progress has been made in the field of education and literacy during the last decades. Still, reading acquisition in SI has as yet not been extensively studied, and as already emphasized several times throughout this thesis, more research is needed to further increase our understanding of reading development and dyslexia in the highly transparent SI language. We therefore hope that the research described in this thesis will turn out to be another step forward in the field of reading research in Indonesia. After all, as even for the most proficient reader-to-be sufficient practice is still the way to go: “*Mari membaca!*”<sup>12</sup>

---

<sup>12</sup> Let's read! (translated from Standard Indonesian).